Numerical Weather Prediction

Error covariance matrices having sinusoidal eigenvectors
I: cyclic domains

Forecasting Research Technical Report No. 446

A. A. White and S. B. Healy

email: nwp_publications@metoffice.com

©Crown Copyright

Error covariance matrices having sinusoidal eigenvectors. I: cyclic domains

A. A. White and S. B. Healy

10th August 2004

Abstract

Covariance matrices of errors - both of forecast and of observation - feature prominently in many data assimilation schemes for numerical weather prediction. To promote understanding of these schemes, we examine situations in which the covariance matrices can be manipulated analytically. If error covariance is homogeneous and symmetric in a cyclic one-dimensional domain, and if grid-points are evenly spaced, then the eigenvectors of the covariance matrix are just sines and cosines whose arguments are linear functions of the grid-point indices. The corresponding eigenvalues are proportional to the Fourier cosine amplitudes of the covariance profile. These general results are well known in various contexts, including data assimilation, but we feel there is much to be gained by elementary consideration of them. The covariance matrix eigenvalue problem can be related to a continuous integral eigenvalue problem which is isomorphic to a simple spatial smoothing problem and whose sinusoidal solutions reflect a specialization of a convolution theorem. For homogeneous, symmetric error covariance, we find sinusoidal solutions of the continuous integral eigenvalue problem in a one-dimensional cyclic domain, with analogous results in two and three dimensions for homogeneous, separably-symmetric error covariance. For homogeneous, axially-symmetric error covariance on the surface of a sphere, the eigen-solutions are spherical harmonics. Insight is gained into the meaning and physical content of error covariance matrices, of the associated eigenvalue problems, of the consequences of uneven spatial discretization, and of various data assimilation procedures. The treatment can readily be extended to the case of homogeneous error correlation coefficients. Part II of this study considers the range of situations in which error covariance (or correlation) matrices have sinusoidal eigenvectors when the spatial domain is non-cyclic or mixed cyclic/non-cyclic.

Contents

Abstract....1

1. Introduction....2

2. 1-dimensional cyclic domain....5

3. 2- and 3-dimensional cyclic domains....11

4. Extension to homogeneous, isotropic covariance on the surface of a sphere....12

5. Concluding remarks....15

Acknowledgements....17

References....17

Appendix: The eigen-problem of the 7×7 matrix (2.5)....19

Figures....20

1 Introduction

Variational data assimilation seeks the best estimate of the atmosphere's state at some chosen time t_0, given a set of observations and a set of *a priori* or "background" values obtained from a previous numerical forecast. In the Bayesian probability framework described by Lorenc (1986), (1992), this best estimate is found by minimising a dimensionless *cost function*, conventionally denoted as J.

J depends on the atmosphere's estimated state $\mathbf{x}$ as a set of independent variables, and on the observations and the background values as sets of parameters. It is the sum of two quadratic forms, J^b and J^o:

$$J(\mathbf{x}) = J^b + J^o, \tag{1.1}$$

where J^b measures the deviation of $\mathbf{x}$ from the background $\mathbf{x}^b$, and J^o measures the deviation of simulated observations $\mathbf{y}$ $(= \mathbf{y}(\mathbf{x}))$ from the real observations $\mathbf{y}^o$. *Covariance matrices* $\mathbf{B}$, $\mathbf{E}$ and $\mathbf{F}$ - soon to be defined - enter the expressions for J^b and J^o :

$$J^b = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) \;; \quad J^o = \frac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T (\mathbf{E}+\mathbf{F})^{-1}\left(\mathbf{y}-\mathbf{y}^o\right). \tag{1.2}$$

Here $\mathbf{x}$, $\mathbf{x}^b$, $\mathbf{y}$ and $\mathbf{y}^o$ are represented as column vectors, and superscript T indicates transposition. (The 1/2 factors will disappear when J is differentiated.) The vectors $\mathbf{x}$ and $\mathbf{x}^b$ contain entries corresponding to each of the numerical model's grid-point values at the chosen initial time t_0. The vectors $\mathbf{y}^o$ and $\mathbf{y}$ (which in general do not have the same number of

components as $\mathbf{x}$ and $\mathbf{x}^b$) contain entries for each observation and its simulated counterpart respectively. $\mathbf{B}$ is the covariance matrix of errors of *forecast* (see (1.6), below), $\mathbf{E}$ is the covariance matrix of errors of *observation*, and $\mathbf{F}$ is the covariance matrix of errors associated with simulating observations from the state $\mathbf{x}$. These covariance matrices arise in the Bayesian estimation process: a Gaussian distribution of errors is assumed, and the sum (1.1) of the quadratic forms (1.2) results when the logarithm is taken - see Lorenc (1986), (1992).

The extremal behaviour of $J(\mathbf{x})$ is easily established when a linear relation $\mathbf{y} = \mathbf{H}\mathbf{x}$ exists between model variables and simulated observations, and $\mathbf{B}$ and $\mathbf{E} + \mathbf{F}$ are independent of $\mathbf{x}$. Differentiation of (1.1) and (1.2) with respect to the components of the vector $\mathbf{x}$ then gives:

$$\frac{\partial J}{\partial \mathbf{x}} = \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right) + \mathbf{H}^T\left(\mathbf{E} + \mathbf{F}\right)^{-1}\left(\mathbf{H}\mathbf{x} - \mathbf{y}^o\right); \tag{1.3}$$

$$\frac{\partial^2 J}{\partial \mathbf{x}^2} = \mathbf{B}^{-1} + \mathbf{H}^T\left(\mathbf{E} + \mathbf{F}\right)^{-1}\mathbf{H}. \tag{1.4}$$

By setting $\partial J/\partial \mathbf{x} = 0$ in (1.3) and noting (1.4), the required solution is obtained as

$$\mathbf{x} - \mathbf{x}^b = -\left(\frac{\partial^2 J}{\partial \mathbf{x}^2}\right)^{-1}\mathbf{H}^T\left(\mathbf{E} + \mathbf{F}\right)^{-1}\left(\mathbf{H}\mathbf{x}^b - \mathbf{y}^o\right). \tag{1.5}$$

In practice, the location of the minimum of J is found by using iterative algorithms; see, for example, Lorenc *et al* (2000), p2997.

The definition of the covariance matrix $\mathbf{B}$ of errors of forecast is

$$\mathbf{B} = \left\langle \left(\mathbf{x}^b - \mathbf{x}^t - \left\langle \mathbf{x}^b - \mathbf{x}^t \right\rangle\right)\left(\mathbf{x}^b - \mathbf{x}^t - \left\langle \mathbf{x}^b - \mathbf{x}^t \right\rangle\right)^T \right\rangle, \tag{1.6}$$

in which the angle brackets $\langle ... \rangle$ represent an ensemble average in the Bayesian sense, and $\mathbf{x}^t$ is the vector of the true values. For brevity, $\mathbf{B}$ may be called the forecast error covariance matrix. (The choice of symbol is conventional, and evidently denotes *background.*) $\mathbf{B}$ contains elements which are the error covariances between every forecast variable in the model. Although vast in size, $\mathbf{B}$ is symmetric. $\mathbf{E}$ and $\mathbf{F}$ are similarly defined in terms of the actual and simulated observation error covariances and are also symmetric matrices.

For definiteness, we have introduced error covariance matrices via a brief discussion of variational data assimilation, but they also arise in other schemes - as does the requirement to manipulate them. See Lorenc (1986) and Daley (1991) for full accounts.

If there were zero covariance between forecast errors in different meteorological fields, and zero covariance between forecast errors in the same field at different places, then $\mathbf{B}$ would be a diagonal matrix of forecast error variances and $\mathbf{B}^{-1}$ would be just the diagonal matrix of their reciprocals. This scenario is useful in a mnemonic sense regarding the form of (1.2), but a key objective of the formalism is to permit a more realistic description of forecast error covariance. The scientific problem is to devise realistic and tractable specifications of $\mathbf{B}$ (Lorenc *et al* 2000, p2998) - a process which is partially guided by the calculated spatial covariances of fields forecast at different lead times but verifying at the same time (Parrish and Derber 1992, Ingleby 2001). A convenient simplification is to work in terms of new fields that are invertibly related to the model's forecast fields $\mathbf{x}$ but whose forecast errors are - at least in a reasonable idealization - uncorrelated (see Lorenc *et al* 2000, p2996).

Then $\mathbf{B}$ is of block diagonal form, with each block corresponding to the spatial covariances between forecast errors in a single field. (In general, the individual blocks are not themselves diagonal.) This simplification is assumed here. We consider a typical individual block of $\mathbf{B}$ and its associated field, but for simplicity we do not introduce a new notation: $\mathbf{B}$ is taken to represent one of the blocks of the forecast error covariance matrix, and $\mathbf{x}$ the associated field. (Of course, if there is only one independent field - for example, in flow governed by the barotropic vorticity equation - then reduction to block diagonal form is not necessary and we can deal directly with the whole forecast error covariance matrix).

In this paper we examine cases in which $\mathbf{B}$ is not diagonal, but can still be manipulated analytically. Such cases are useful in many ways. They help to develop the conceptual and mathematical framework for understanding what must be done, and how it can be done, in more complicated situations. They sharpen appreciation of the physical significance of $\mathbf{B}$ and of its eigenvectors and eigenvalues. We shall exploit an analogy between the covariance eigenvalue problem and a smoothing eigenvalue problem (which aids physical understanding) and demonstrate the relevance of *integral equation eigenvalue problems* (which clarifies discretization and weighting issues, and exposes the mathematical foundations). A very advantageous feature, which is implicit in the work of Boer (1983), Courtier *et al.* (1998), Gaspari and Cohn (1999) and others, is that the range of cases susceptible to simple analytical treatment is wide: any case in which the error covariance is homogeneous and separably-symmetric in a cyclic Cartesian domain may be treated, and any case in which the error covariance is homogeneous and axially-symmetric on the surface of a sphere. In some non-cyclic domains, and some mixed domains (such as the 3-dimensional spherical annulus domain of the Earth's atmosphere), the situation is less favourable - as we shall see in Part II (White and Healy 2004). Overall, however, the range of cases in which sinusoidal eigenvectors occur is sufficiently wide that their use in practical data assimilation becomes a realistic proposition if we are uncertain about the best detailed form of covariance matrix elements (or, in some instances, if we are content to approximate a known form). It becomes an attractive proposition if algorithmic acceleration may be achieved by use of Fourier Transforms. Some of these aspects are illustrated in Healy and White (2004) in the context of observation error covariance - for which Rodgers (1990) has used convenient analytical profiles.

In Section 2 we examine the archetypal case of homogeneous, symmetric error covariance of a single field in a cyclic, 1-dimensional domain. We show that the eigenvectors are simple sinusoids, and derive results involving the Fourier transform of the covariance profile, a convolution theorem and the underlying integral structure. Section 3 notes extensions to cyclic domains in 2 and 3 dimensions. Section 4 derives parallel results for a (2D) spherical surface domain. Insights gained are summarised in the concluding section 5, and some implications noted.

This paper does not apply systematic methods for solving integral equations or assume a knowledge of them. Although we regard the underlying integral structure as central to an appreciation of covariance matrices, we shall simply *observe* that certain integral equations have sinusoidal solutions (as is readily verified by substitution).

2 1-dimensional cyclic domain

Suppose that grid-points $x = x_j,\ j = 1, 2, ...N + 1$ are equally spaced in a cyclic domain of length $2L$, with $x_j = 2(j-1)L/N$; then $F(x_{N+1}) = F(x_1)$ for a given meteorological field F. See Figure 1. The grid-points might, for example, be spaced at equal intervals around a latitude circle. Suppose also that the covariance of forecast errors in F depends only on the distance between the relevant grid-points (i.e. that the covariance is homogeneous and symmetric). The error covariance matrix $\mathbf{B}$ will then not only be symmetric, it will be cyclic in a well-defined sense; of its N^2 elements, only $(N+1)/2$ will be independent if N is odd, and $N/2$ if N is even (of which one may be chosen as unit). We assume for clarity that N is odd. (The case of even N is more frequently encountered in practice, but is at least aesthetically unsatisfactory because an odd number of amplitudes is needed to specify a uniformly truncated full-range Fourier series).

Consider the first row of the error covariance matrix $\mathbf{B}$. The elements B_{1k}, $(k = 1, 2, ...N)$ represent the covariance between the error at grid-point 1 and the error at grid-point k. Since the error covariance is symmetric, B_{12} must be equal to B_{1N} because grid-point N is the same distance from grid-point 1 as is grid-point 2 [see Figure 1]. There will be $(N-1)/2$ equalities of this sort, and hence only $(N+1)/2$ independent elements in the first row of $\mathbf{B}$.

For example, if $N = 7$ then the first row of $\mathbf{B}$ will have the form

$$(B_{11}\ \ B_{12}\ \ B_{13}\ \ B_{14}\ \ B_{14}\ \ B_{13}\ \ B_{12})\,. \tag{2.1}$$

It is convenient to choose B_{11} as unit, and to define

$$a \equiv B_{12}\ ,\ \ b \equiv B_{13}\ ,\ \ c \equiv B_{14}\,. \tag{2.2}$$

The first row of $\mathbf{B}$ now takes a visually less demanding form:

$$(1\ \ a\ \ b\ \ c\ \ c\ \ b\ \ a)\,. \tag{2.3}$$

The second row of $\mathbf{B}$ represents the covariance betwen the error at grid-point 2 and the error at grid-point k. Since the error covariance is homogeneous, these covariances are equal to those between the errors at grid-point 1 and grid-point $k-1$ (the grid-points being equally spaced). Hence the second row of $\mathbf{B}$ is a one-step cyclic permutation of the first row (2.3) of $\mathbf{B}$:

$$(a\ \ 1\ \ a\ \ b\ \ c\ \ c\ \ b)\,. \tag{2.4}$$

The whole matrix (in this $N = 7$ case) is the following symmetric, cyclic form:

$$\mathbf{B} = \begin{pmatrix} 1 & a & b & c & c & b & a \\ a & 1 & a & b & c & c & b \\ b & a & 1 & a & b & c & c \\ c & b & a & 1 & a & b & c \\ c & c & b & a & 1 & a & b \\ b & c & c & b & a & 1 & a \\ a & b & c & c & b & a & 1 \end{pmatrix} \tag{2.5}$$

$\mathbf{B}$ has only 3 independent elements (another having been chosen as unit). Each row is a cyclic one-step (right shift) permutation of its predecessor, and there is symmetry about

the leading diagonal. Cyclic matrices are also known as circulant matrices. Because all the elements along a given diagonal are equal, $\mathbf{B}$ is also of symmetric Toeplitz form (see Part II) but the symmetric cyclic property is more important in the present discussion.

It is well known that symmetric circulant matrices possess sinusoidal eigenvectors. For example, in a study of time-series covariance, Preisendorfer (1988) [see pp129-146] applies some general matrix theorems and the properties of a simple permutation matrix to show that the eigenvectors and eigenvalues of such matrices may be determined analytically, and (in particular) that the eigenvectors are sinusoidal. Here we shall use a different treatment which leads to the same results and indicates the connection of the discrete eigenvalue problem for the matrix $\mathbf{B}$ to a continuous eigenvalue problem for a related integral operator. This connection proves to be helpful in several ways.

One eigenvector of the 7×7 form (2.5), and its eigenvalue, may be found by inspection. Each row of the matrix has the same sum $(1 + 2a + 2b + 2c)$, so each row gives the same result when it operates on a column vector having identical elements. Hence such a column vector is an eigenvector of the 7×7 form (2.5), and the corresponding eigenvalue is simply $(1 + 2a + 2b + 2c)$. The normalised eigenvector is $\left(\sqrt{7}\right)^{-1} (1, 1, 1, 1, 1, 1, 1)^T$.

The other eigenvectors may be found by noting that each row of $\mathbf{B}$ has the same Fourier wavenumber spectrum in terms of sinusoids on the 7-point grid, but with a phase shift which is associated with a change of sine and cosine contributions from row to row. Analytical treatment (see Appendix) confirms an intuitive expectation (see Figure 2) that the other eigenvectors are simply the three sines and three (non-uniform) cosines that are periodic on the 7-point grid; also, the corresponding eigenvalues *are proportional to the amplitudes of the cosine components in the Fourier decomposition of the symmetric row profile* (as it appears, for example, in the fourth row of $\mathbf{B}$). The eigenvector $\left(\sqrt{7}\right)^{-1} (1, 1, 1, 1, 1, 1, 1)^T$ fits into this picture, since its structure is that of the zero-wavenumber cosine, and the corresponding eigenvalue $(1 + 2a + 2b + 2c)$ is $7\times$ the mean value of the row profile of $\mathbf{B}$.

These results can be generalised to arbitrary N. But discrete Fourier expansions may be avoided, and the underlying physical situation displayed in its simplicity, by considering the limit in which each "row" of $\mathbf{B}$ is infinite in length and may be represented as a continuous Fourier series. This corresponds to the transition from discrete covariance between errors at different grid-points to continuous covariance between errors at different continuously-varying locations. Under the assumed conditions of homogeneity and symmetry of error covariance, the typical element B_{jk} of the discrete covariance matrix is a function only of the separation of points x_j and x_k. We may therefore write

$$B_{jk} = S\left(x_k - x_j\right), \tag{2.6}$$

where S is a symmetric function of its argument and represents the generic row "profile" of matrix $\mathbf{B}$. In the continuous limit $S\left(x_k - x_j\right)$ becomes $S\left(x - X\right)$ where x and X are continuously-varying coordinates. Instead of dealing with sums over the repeated index k when considering the action of matrix $\mathbf{B}$ on any column vector $\mathbf{v}$, i.e. $\sum B_{jk} v_k$, in the continuous case we deal with integrals over x when considering the inner product of $S\left(x - X\right)$ and the function $v(x)$, i.e. $\int S\left(x - X\right) v(x) dx$. S is a kernel function in such integrals.

Each "row" in the continuous limit has the same Fourier spectrum in terms of wavenum-

ber, but a different phase. The generic "row" may be expressed as $S(x-X)$ where

$$S(x-X) = S_0 + \sum_{m=1}^{\infty} 2S_m \cos\left(\pi m\left(x-X\right)/L\right). \tag{2.7}$$

Here S_0 is the average value of S over $0 \le x \le 2L$, the Fourier cosine series expansion ensures symmetry of S about the centre X of the profile, and the factor 2 has been introduced for early numerical convenience. Relative to X as origin, the profile of S is

$$S(x) = S_0 + \sum_{m=1}^{\infty} 2S_m \cos\left(\pi mx/L\right) = \sum_{m=-\infty}^{m=+\infty} S_m \exp\left(i\pi mx/L\right). \tag{2.8}$$

This is the essential "shape" of S; in other "rows" ($X \neq 0$) the profile is displaced as indicated by (2.7). In (2.8), $i = \sqrt{-1}$ and the (real) coefficients for $m < 0$ are given by $S_{-m} = S_m$.

Define a domain wavenumber α as

$$\alpha = \frac{\pi}{L}. \tag{2.9}$$

Upon use of a standard trig identity, (2.7) then becomes

$$S(x-X) = S_0 + \sum_{m=1}^{\infty} 2S_m \left[\cos\alpha mX \cos\alpha mx + \sin\alpha mX \sin\alpha mx\right]. \tag{2.10}$$

By multiplying both sides of (2.10) by $\cos\alpha nx$, integrating over $x = [0, 2L]$, exploiting well-known orthogonality properties of sines and cosines, and renaming n as m, one finds

$$\frac{1}{2L}\int_0^{2L} S(x-X)\cos\alpha mx dx = S_m \cos\alpha mX; \quad 0 \le m \le \infty. \tag{2.11}$$

A similar procedure using $\sin\alpha nx$ leads to

$$\frac{1}{2L}\int_0^{2L} S(x-X)\sin\alpha mx dx = S_m \sin\alpha mX; \quad 1 \le m \le \infty. \tag{2.12}$$

Consider, on the cyclic interval $[0, 2L]$, the *integral eigenvalue problem*

$$\frac{1}{2L}\int_0^{2L} S(x-X)f(x)dx = \mu f(X), \tag{2.13}$$

in which eigenfunctions $f = f_m$ and associated eigenvalues $\mu = \mu_m$ are sought. Results (2.11) and (2.12) reveal solutions to this problem. *Eigenfunctions* $f_m(x)$ of (2.13) *are the familiar sines and cosines* $\sin\alpha mx$ *and* $\cos\alpha mx$ *on the cyclic interval* $[0, 2L]$ (recall $\alpha = \pi/L$) *and the corresponding eigenvalues* μ_m *are equal to the coefficients* S_m *in the Fourier cosine expansion* (2.8) *of the symmetric kernel* $S(x)$. In this sense, the eigenvalues are the Fourier cosine transform of the symmetric kernel $S(x)$ - an interpretation discussed further in Part II in the context of infinite, non-cyclic domains.

Several aspects and consequences are worth noting and exploring. We denote by $\{\sin\alpha mx\}$ and $\{\cos\alpha mx\}$ the classes of functions $\sin\alpha mx$ having $m \ge 1$ and $\cos\alpha mx$ having $m \ge 0$.

1. The eigenfunctions are the same simple sines and cosines $\{\sin \alpha mx\}$ and $\{\cos \alpha mx\}$ for *all* symmetric kernels S.

2. The symmetric kernel S determines the eigenvalues: they are the coefficients appearing in the Fourier cosine spectrum of S - see (2.8), (2.11) and (2.12).

3. Since, for each m, the eigenfunctions $\sin \alpha mx$ and $\cos \alpha mx$ have the same eigenvalue, the eigen-structure of the problem is fully pair-wise degenerate except for the case $m = 0$ (which corresponds to the equal-element eigenvector of (2.5) previously noted).

4. The kernel S in (2.11) and (2.12) and in the integral equation (2.13) could be regarded as a symmetric weighting or smoothing applied to the input function $f(x)$. Consider the integral on the left-hand side of (2.13) as the *definition* of a new function $g(X)$ (for given choices of f and S):

$$g(X) = \frac{1}{2L} \int_0^{2L} S(x - X) f(x) dx. \tag{2.14}$$

Because both f and S are cyclic functions with spatial period $2L$, the integral in (2.14) will have the same value if the limits are changed from $x = 0$ and $x = 2L$ to $x = c$ and $x = 2L + c$ for any real c; see Figure 3. Choose $c = X - L$; then (2.14) becomes

$$g(X) = \frac{1}{2L} \int_0^{2L} S(x - X) f(x) dx = \frac{1}{2L} \int_{X-L}^{X+L} S(x - X) f(x) dx. \tag{2.15}$$

Introducing the new variable $\widehat{x} = x - X$ (noting that $d\widehat{x} = dx$) then shows that

$$g(X) = \frac{1}{2L} \int_{-L}^{L} S(\widehat{x}) f(X + \widehat{x}) d\widehat{x} = \frac{1}{2L} \int_{-L}^{L} S(x) f(X + x) dx. \tag{2.16}$$

[Replacing $\widehat{x}$ by x in (2.16) is permissible because both are dummy variables of integration.] From (2.16) it is clear that $g(X)$ is the result of applying S as a symmetric weighting to f over an interval of width $2L$ centred on X; see Figure 4. We may imagine g being generated from f by applying the weight function S centrally to every point in our cyclic domain. This in itself is a useful view, but more important is the expectation - based on experience of simple smoothing methods - that *if f is any sinusoid that obeys the periodicity conditions, then g will have the same form but a different amplitude.* So any sine or cosine that obeys the periodicity conditions is expected to be an eigenfunction of the integral problem

$$\frac{1}{2L} \int_{-L}^{L} S(x) f(X + x) dx = \frac{1}{2L} \int_0^{2L} S(x - X) f(x) dx = \mu f(X), \tag{2.17}$$

the eigenvalue μ being a measure of the ratio of amplitudes of the output and input sinusoids. We have confirmed this expectation by the analysis leading to (2.11) and (2.12), which has also shown that the eigenvalues are the coefficients in the Fourier cosine expansion of the symmetric weighting function S.

5. The smoothing analogy can be viewed within the context of a *convolution theorem* for full-range Fourier series. Suppose that f and g each have a standard Fourier expansion of the form

$$F(x) = \sum_{m=-\infty}^{m=+\infty} F_m \exp(i\alpha mx) \tag{2.18}$$

in which $F_{-m} = (F_m)^*$, where * indicates complex conjugation. The second equality in (2.8) gives a similar expansion for $S(x)$ (with real coefficients S_m, since S is symmetric). As

before, $\alpha = \pi/L$, and f, g and S have periodicity $2L$. The inversion rule for the F_m in (2.18) is

$$F_m = \frac{1}{2L}\int_{-L}^{L} F(x)\exp\left(-i\alpha m x\right)dx. \tag{2.19}$$

By applying (2.19) to the expression (2.16) for g in terms of S and f, it follows that

$$g_m = \frac{1}{4L^2}\int_{-L}^{L}\exp\left(-i\alpha m x\right)\left\{\int_{-L}^{L} S(x')f(x+x')dx'\right\}dx. \tag{2.20}$$

By successively (i) reversing the order of integration, (ii) defining a new variable $x'' = x + x'$, (iii) noting that additions c may be made to the limits in the integrals without changing their values, (iv) noting the symmetry of S (i.e. $S(x) = S(-x)$) and (v) re-labelling dummy variables, (2.20) may be reduced to

$$g_m = \frac{1}{4L^2}\int_{-L}^{L} S(x)\exp\left(-i\alpha m x\right)dx\int_{-L}^{L} f(x)\exp\left(-i\alpha m x\right)dx = S_m f_m. \tag{2.21}$$

Each Fourier coefficient of the resultant smoothed function is equal to the product of the corresponding Fourier coefficient of the function being smoothed and the corresponding Fourier coefficient of the kernel function that specifies the smoothing. So if the function being smoothed is sinusoidal (with periodicity $2L$) then the result of the smoothing is the same sinusoid, but with amplitude modified by the factor S_m. This is the same as our earlier conclusion, and it is obvious once the convolution theorem (2.21) is established. However, the labour of establishing (2.21) is much greater than that of our earlier direct demonstration of the conclusion. [If the domain is not cyclic then the convolution theorem must be reconsidered.]

6. The relationship between integral eigenvalue problems and matrix eigenvalue problems is discussed in textbooks such as Courant and Hilbert (1953) and Mathews and Walker (1965). The integral eigenvalue problem (2.13) may be discretized on a grid of points x_j, perhaps unequally spaced, as (recall (2.6))

$$\frac{1}{2L}\sum_k S\left(x_k - x_j\right) f\left(x_k\right)\delta x_k = \frac{1}{2L}\sum_k B_{jk} f\left(x_k\right)\delta x_k = \mu f\left(x_j\right), \tag{2.22}$$

where $\delta x_k = \left(x_{k+1} - x_{k-1}\right)/2$. If $\mathbf{D}$ is the diagonal matrix of $\delta x_k/2L$, (2.22) may be written

$$\mathbf{BDf} = \mu\mathbf{f}, \tag{2.23}$$

where $\mathbf{f}$ is the column vector whose kth element is $f_k = f\left(x_k\right)$. If the N grid-points are evenly spaced at intervals Δx (as shown in Figure 1), then $\mathbf{D}$ is a scalar multiple of the identity matrix $\mathbf{I}$, and (2.23) reduces to

$$\mathbf{Bf} = \left(\frac{2\mu L}{\Delta x}\right)\mathbf{f} = \mu N\mathbf{f}, \tag{2.24}$$

which is the eigenvalue problem for the symmetric, circulant covariance matrix $\mathbf{B}$ with each element divided by N. The eigenfunctions are sinusoids evaluated at the grid-points; see the Appendix for the case $N = 7$. If the spacing is uneven, (2.23) involves a non-symmetric, non-circulant matrix $\mathbf{BD}$, but this merely reflects the uneven spatial sampling: a simple

situation in physical space may be inconveniently distorted in the "index space" of unevenly spaced grid-points. Equation (2.23) may be re-written as

$$\mathbf{D}^{1/2}\mathbf{B}\mathbf{D}^{1/2}\left(\mathbf{D}^{1/2}\mathbf{f}\right) = \mu\left(\mathbf{D}^{1/2}\mathbf{f}\right). \tag{2.25}$$

Thus $\mathbf{D}^{1/2}\mathbf{f}$ are eigenvectors of the *symmetric* matrix $\mathbf{D}^{1/2}\mathbf{B}\mathbf{D}^{1/2}$ with the same eigenvalues μ as those of the eigenvectors $\mathbf{f}$ of (2.23). (Since $\mathbf{D}$ is diagonal, $\mathbf{D}^{1/2}$ is also diagonal and is readily calculated.) It follows that the eigenvectors $\mathbf{f}$ of (2.23) are orthogonal with respect to weights that are the diagonal elements of $\mathbf{D}$. To a good conceptual approximation, the eigenvectors $\mathbf{f}$ of (2.23) will be sinusoids in x evaluated at the locations of the grid-points.

7. Suppose that the grid-points are evenly spaced. Having found the eigenvectors and eigenvalues of the circulant matrix $\mathbf{B}$, its inverse may be obtained by standard procedures for symmetric matrices. If $\mathbf{M}$ is the matrix whose columns are the normalised eigenvectors ($\mathbf{f}$) of $\mathbf{B}$, and $\mathbf{L}$ is the diagonal matrix of its eigenvalues (suitably ordered) then

$$\mathbf{M}^T\mathbf{B}\mathbf{M} = \mathbf{L}. \tag{2.26}$$

(The eigenvectors of $\mathbf{B}$ are orthogonal because of its symmetry.) Hence, in terms of the diagonal matrix $\mathbf{L}^{-1}$ of the reciprocals of the eigenvalues of $\mathbf{B}$, we find

$$\mathbf{B}^{-1} = \mathbf{M}\mathbf{L}^{-1}\mathbf{M}^T. \tag{2.27}$$

In the present context, the most useful aspect of this formal result is the scope offered for simplification of the background error part $\left(J^b\right)$ of the expression (1.2) for $J(\mathbf{x})$. In terms of the vector $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$ of *increments* we have:

$$\left(\mathbf{x}-\mathbf{x}^b\right)^T\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) = (\Delta\mathbf{x})^T\mathbf{M}\mathbf{L}^{-1}\mathbf{M}^T\Delta\mathbf{x} = \left(\mathbf{M}^T\Delta\mathbf{x}\right)^T\mathbf{L}^{-1}\left(\mathbf{M}^T\Delta\mathbf{x}\right). \tag{2.28}$$

Now

$$\widetilde{\Delta\mathbf{x}} = \mathbf{M}^T\Delta\mathbf{x} \tag{2.29}$$

is the projection of $\Delta\mathbf{x}$ onto the (sinusoidal) eigenvectors of $\mathbf{B}$. In other words, $\widetilde{\Delta\mathbf{x}}$ is the Fourier spectrum of $\Delta\mathbf{x}$. In terms of $\widetilde{\Delta\mathbf{x}}$, (2.28) becomes

$$\left(\mathbf{x}-\mathbf{x}^b\right)^T\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) = \left(\widetilde{\Delta\mathbf{x}}\right)^T\mathbf{L}^{-1}\widetilde{\Delta\mathbf{x}}. \tag{2.30}$$

Moreover, in terms of

$$\widehat{\Delta\mathbf{x}} = \mathbf{L}^{-1/2}\mathbf{M}^T\Delta\mathbf{x}, \tag{2.31}$$

(2.28) reduces to

$$\left(\mathbf{x}-\mathbf{x}^b\right)^T\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) = \left(\widehat{\Delta\mathbf{x}}\right)^T\widehat{\Delta\mathbf{x}}. \tag{2.32}$$

$\mathbf{L}^{-1/2}$ is the diagonal matrix of the reciprocals of the square roots of the eigenvalues of $\mathbf{B}$. $\widehat{\Delta\mathbf{x}}$ is a scaled projection of the increment vector onto the (sinusoidal) eigenvectors of $\mathbf{B}$, the scaling of each term involving the reciprocal of the square root of the associated eigenvalue. A major benefit of the simplifying transformation (2.31) is that the conditioning has been simultaneously improved (see Lorenc *et al.* 2000, p2994). However, cases in which $\mathbf{B}$ has one zero eigenvalue (or more) remain unreformed: in such cases - which are readily constructed,

since the eigenvalues are just the Fourier amplitudes of the covariance profile - one (or more) of the diagonal terms of $\mathbf{L}$ is zero, so $\mathbf{L}$ cannot be inverted. Neither can negative-eigenvalue cases be treated.

8. When the grid-points are unevenly spaced, a similar analysis may be applied via the symmetric matrix $\mathbf{D}^{1/2}\mathbf{B}\mathbf{D}^{1/2}$ and its eigenfunctions $\mathbf{g} = \mathbf{D}^{1/2}\mathbf{f}$ (see (2.25)). If $\mathbf{G}$ is the matrix whose columns are the normalised eigenfunctions $\mathbf{g}$, then in place of (2.26) we have

$$\mathbf{G}^T\mathbf{D}^{1/2}\mathbf{B}\mathbf{D}^{1/2}\mathbf{G} = \mathbf{L}\,. \tag{2.33}$$

Hence

$$\mathbf{B}^{-1} = \left(\mathbf{D}^{1/2}\mathbf{G}\right)\mathbf{L}^{-1}\left(\mathbf{D}^{1/2}\mathbf{G}\right)^T = (\mathbf{D}\mathbf{M})\,\mathbf{L}^{-1}\,(\mathbf{D}\mathbf{M})^T\,. \tag{2.34}$$

Thus (2.30) is achieved if now

$$\widetilde{\Delta\mathbf{x}} = \mathbf{M}^T\mathbf{D}\Delta\mathbf{x}, \tag{2.35}$$

and (2.32) if

$$\widehat{\Delta\mathbf{x}} = \mathbf{L}^{-1/2}\mathbf{M}^T\mathbf{D}\Delta\mathbf{x}. \tag{2.36}$$

Here $\mathbf{M}$ is the matrix whose columns are the eigenvectors of the non-symmetric matrix $\mathbf{B}\mathbf{D}$; see (2.23) and the ensuing discussion.

9. The range of eigenvalues of the covariance matrix $\mathbf{B}$ (which determines the conditioning of the untransformed variational analysis problem) is determined by the projection of its discrete row profile on the cosines $\{\cos\alpha mx\}$ evaluated at the grid-points. [There is no projection onto the sines $\{\sin\alpha mx\}$ because the profile is symmetric.] Two extreme cases are of particular interest. If the row profile has all zero elements except for an isolated central peak, then each Fourier component has equal amplitude and so the eigenvalues are identical; this corresponds to a well-posed case in which $\mathbf{B}$ is diagonal. If the row profile of $\mathbf{B}$ has a Fourier decomposition with zero projection on any member of $\{\cos\alpha mx\}$ within the truncation, then $\mathbf{B}$ has a zero eigenvalue and the variational analysis problem is ill-posed (because $\mathbf{B}^{-1}$ will have a corresponding infinite eigenvalue and the condition number will also be infinite; see Wilkinson (1964), section 4.2).

3 2- and 3-dimensional cyclic domains

The arguments given in section 2 are readily extended to cyclic domains in two Cartesian dimensions. Suppose that the domain scales in x and y are $2L_x$ and $2L_y$ (see Figure 5), and define domain wavenumbers α and β as

$$\alpha = \frac{\pi}{L_x}, \qquad \beta = \frac{\pi}{L_y}. \tag{3.1}$$

Suppose also that the covariance kernel $S(x,\,y)$ is *separably symmetric* in the sense that

$$S(x,\,y) = S_1(x)S_2(y) \tag{3.2}$$

where both S_1 and S_2 are symmetric functions of their arguments, and periodic on the domain. [This stipulation means that $S(x,\,y)$ is *centrally* symmetric, i.e. $S(-x,\,-y) = S(x,\,y)$, but not *radially* symmetric. Hence the error covariance, though homogeneous, is

not isotropic. In one periodic Cartesian dimension, symmetry implies isotropy; in two or more periodic Cartesian dimensions radial symmetry (and hence isotropy) is possible only in the trivial case of uniform covariance. In the compact domain of a spherical surface, non-trivial axial symmetry and hence isotropy are of course allowed. See section 4.] Then the eigenfunctions f of the integral eigenvalue problem

$$\frac{1}{4L_xL_y}\int_0^{2L_x}\int_0^{2L_y} S(x-X,\ y-Y)f(x,y)dxdy = \mu f(X,Y) \tag{3.3}$$

are products of sines and cosines: $\sin\alpha mx \sin\beta ny$, $\sin\alpha mx\cos\beta ny$, $\cos\alpha mx\cos\beta ny$ and $\cos\alpha mx\sin\beta ny$. Further, the corresponding eigenvalues μ are proportional to the coefficients of $\cos\alpha mx\cos\beta ny$ in the Fourier decomposition of the (periodic) covariance function $S(x,y)$. [For given m and n, the eigenvalues are the same for each of these four products, so the problem is 4-fold degenerate. Alternative eigenfunctions, with the same eigenvalues, are therefore $\sin(\alpha mx \pm \beta ny)$ and $\cos(\alpha mx \pm \beta ny)$.] These results may be demonstrated by the direct method used in section 2 for the 1D case. They may be related to known or intuitive results about smoothing by applying periodicity arguments to the integral in (3.3) to show that

$$\int_0^{2L_x}\int_0^{2L_y} S(x-X,\ y-Y)f(x,y)dxdy = \int_{-L_x}^{L_x}\int_{-L_y}^{L_y} S(x,\ y)f(X+x,Y+y)dxdy. \tag{3.4}$$

Hence (3.3) is equivalent to

$$\frac{1}{4L_xL_y}\int_{-L_x}^{L_x}\int_{-L_y}^{L_y} S(x,\ y)f(X+x,Y+y)dxdy = \mu f(X,Y). \tag{3.5}$$

The eigenfunctions of (3.5) are those functions whose *form* remains unchanged under the smoothing operation represented by the integral on the left side; see Figure 5. These functions are the products of sines and cosines previously noted. The eigenvalues correspond to the amplitude factors of the smoothing, and are equal to the coefficients of $\cos\alpha mx\cos\beta ny$ in the Fourier decomposition of $S(x,y)$. [Multiple zero eigenvalues are possible; consider, for example, the case $S(x,y) = 4\cos^2\alpha x\cos^2\beta y = 1+\cos 2\alpha x+\cos 2\beta y+\cos 2\alpha x\cos 2\beta y$.]

These results extend in the expected way to cyclic domains in 3 Cartesian dimensions. In Part II (White and Healy 2004), we shall find that sinusoidal eigenfunctions abound in infinite domains, but that other non-cyclic and mixed cyclic/non-cyclic domains are far less supportive of them. Irrespective of the occurrence or non-occurrence of sinusoidal eigenfunctions, however, and in both 2 and 3 dimensions, discretization of the continuous integrals (such as (3.3) and (3.5) in 2D) can be helpful when visualising the structures of the corresponding covariance matrices.

4 Extension to homogeneous, isotropic covariance on the surface of a sphere

Consider the integral eigenvalue problem (for eigenfunctions f and eigenvalues μ),

$$\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} S(\lambda,\phi;\Lambda,\Phi)f(\lambda,\phi)\cos\phi d\phi d\lambda = \mu f(\Lambda,\Phi), \tag{4.1}$$

in which the kernel $S(\lambda,\phi;\Lambda,\Phi)$ depends only on the distance between the variable point (λ,ϕ) and the current point (Λ,Φ), λ and Λ being longitudes, and ϕ and Φ latitudes; see Figure 6. This kernel corresponds to homogeneous, isotropic error covariance on the sphere. Since $S(\lambda,\phi;\Lambda,\Phi)$ is symmetric about the axis (Λ,Φ), it must be expressible as a linear combination of ordinary Legendre polynomials P_l (i.e. zonal harmonics) referred to the axis (Λ,Φ):

$$S=\sum_{l=0}^{l=\infty} S_l P_l(\cos\gamma). \tag{4.2}$$

Here γ is the angle subtended by points (λ,ϕ) and (Λ,Φ) - in other words, the co-latitude of (λ,ϕ) relative to (Λ,Φ) as pole; see Figure 6.

Eigenfunctions of (4.1) may be found experimentally. Suppose that the function $f(\lambda,\phi)$ inside the integral on the left-hand side of (4.1) is a (surface) spherical harmonic $Y_n^m(\lambda,\phi)$ referred to the usual (polar) axis: $Y_n^m(\lambda,\phi)=P_n^{|m|}(\sin\phi)\,e^{im\lambda}$, where P_n^m is the associated Legendre function (of the first kind) of degree n and rank m. An invariance property of surface spherical harmonics (reflecting their status as eigenfunctions of the 2D Laplacian operator on the surface of a sphere, independent of the adopted coordinate system) implies that $f(\lambda,\phi)=Y_n^m(\lambda,\phi)$ must be expressible as a linear combination of spherical harmonics of the same degree n but referred to the axis (Λ,Φ) rather than the usual polar axis. See, for example, Sardeshmukh and Hoskins (1984), p2525. Upon substitution of the relevant series into the left-hand side of (4.1) and use of (4.2) and standard spherical harmonic orthogonality properties, it is found that *only the term representing projection onto the P_n component of S survives*:

$$\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} S(\lambda,\phi;\Lambda,\Phi)Y_n^m(\lambda,\phi)\cos\phi\,d\phi\,d\lambda=\left(\frac{4\pi}{2n+1}\right)A_n^0 S_n. \tag{4.3}$$

In (4.3), A_n^0 is the coefficient of P_n in the expression of $Y_n^m(\lambda,\phi)$ as a linear combination of surface spherical harmonics referred to the axis (Λ,Φ). The argument leading to (4.3) applies that given by Ramsey (1949), section 6.44. Ramsey then notes that A_n^0 is equal to the value of $Y_n^m(\lambda,\phi)$ at the point (Λ,Φ), since all the tesseral (i.e. non-zonally symmetric) elements in the expansion of $Y_n^m(\lambda,\phi)$ as spherical harmonics of degree n referred to the axis (Λ,Φ) vanish there! Hence (4.3) may be written

$$\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} S(\lambda,\phi;\Lambda,\Phi)Y_n^m(\lambda,\phi)\cos\phi\,d\phi\,d\lambda=\left(\frac{4\pi}{2n+1}\right)S_n Y_n^m(\Lambda,\Phi). \tag{4.4}$$

A spherical harmonic Y_n^m is an eigenfunction of (4.1)*, with eigenvalue $4\pi S_n/(2n+1)$.* (The assumed symmetry of $S(\lambda,\phi;\Lambda,\Phi)$ about the axis (Λ,Φ) should be emphasised. Note also that the normalization used here for spherical harmonics is to a value of unity at the North Pole, rather than to unit variance over the sphere.)

The integral eigenvalue problem (4.1) may be written more concisely as

$$\int S(\omega-\Omega)f(\omega)d\omega=\mu f(\Omega). \tag{4.5}$$

Here ω and Ω indicate the latitude and longitude combinations (λ,ϕ) and (Λ,Φ) that appear explicitly in (4.1), $d\omega$ is an area element on the sphere, and $S(\omega-\Omega)$ is understood to be

axisymmetric - that is, to depend only on the angle γ between the two directions specified by ω and Ω. See, for example, Mathews and Walker (1965), p169.

The eigenvalue problem (4.1) is isomorphic to a smoothing problem. The functions that are unchanged in *form* by any smoothing that is axially symmetric about the direction of the current point (Λ, Φ) are the eigenfunctions of (4.1). They are spherical harmonics. Each eigenvalue is determined by the projection of the symmetric kernel S onto the zonal harmonic of the same degree (n) as the corresponding spherical harmonic eigenfunction (see (4.4)).

Analogues of the discretization results derived in section 2 for the 1D case may be derived. In terms of a (diagonal) matrix $\mathbf{A}$ of areas, the discrete eigenvalue problem is of the form

$$\mathbf{BAf} = \mu\mathbf{f}. \tag{4.6}$$

If grid-points are evenly spaced both in longitude λ and in latitude ϕ, then $\mathbf{A}$ is not a scalar multiple of the unit matrix, and $\mathbf{BA}$ is not symmetric. As in the 1D case, this reflects non-uniform spatial sampling rather than any intrinsic complication in physical space; the eigenvectors of $\mathbf{BA}$ are expected to be well approximated by spherical harmonics evaluated at the grid-points. From (4.6), the symmetric matrix $\mathbf{A}^{1/2}\mathbf{BA}^{1/2}$ has eigenvectors $\mathbf{g} = \mathbf{A}^{1/2}\mathbf{f}$, where $\mathbf{f}$ is any eigenvector of $\mathbf{BA}$, and $\mathbf{g}$ and $\mathbf{f}$ have the same eigenvalues. If $\mathbf{G}$ is the matrix whose columns are the normalised eigenvectors $\mathbf{g}$, and $\mathbf{L}$ is the diagonal matrix of the corresponding eigenvalues, then

$$\mathbf{G}^T\mathbf{A}^{1/2}\mathbf{BA}^{1/2}\mathbf{G} = \mathbf{L}\,. \tag{4.7}$$

Hence

$$\mathbf{B}^{-1} = \left(\mathbf{A}^{1/2}\mathbf{G}\right)\mathbf{L}^{-1}\left(\mathbf{A}^{1/2}\mathbf{G}\right)^T = (\mathbf{AM})\,\mathbf{L}^{-1}\,(\mathbf{AM})^T\,, \tag{4.8}$$

in which $\mathbf{M}$ is the matrix whose columns are the normalised (approximately spherical harmonic) eigenvectors of $\mathbf{BA}$. In terms of the diagonal matrix $\mathbf{L}^{-1}$ and the vector $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$ of increments,

$$\left(\mathbf{x}-\mathbf{x}^b\right)^T\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) = (\Delta\mathbf{x})^T\,\mathbf{AML}^{-1}(\mathbf{AM})^T\Delta\mathbf{x} = \left(\mathbf{M}^T\mathbf{A}\Delta\mathbf{x}\right)^T\mathbf{L}^{-1}\left(\mathbf{M}^T\mathbf{A}\Delta\mathbf{x}\right). \tag{4.9}$$

With

$$\widetilde{\Delta\mathbf{x}} = \mathbf{M}^T\mathbf{A}\Delta\mathbf{x}, \tag{4.10}$$

which is the projection of $\Delta\mathbf{x}$ onto the eigenvectors of $\mathbf{BA}$, (4.9) becomes

$$\left(\mathbf{x}-\mathbf{x}^b\right)^T\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) = \left(\widetilde{\Delta\mathbf{x}}\right)^T\mathbf{L}^{-1}\widetilde{\Delta\mathbf{x}}. \tag{4.11}$$

Also, in terms of

$$\widehat{\Delta\mathbf{x}} = \mathbf{L}^{-1/2}\mathbf{M}^T\mathbf{A}\Delta\mathbf{x}, \tag{4.12}$$

(4.9) reduces to

$$\left(\mathbf{x}-\mathbf{x}^b\right)^T\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) = \left(\widehat{\Delta\mathbf{x}}\right)^T\widehat{\Delta\mathbf{x}}. \tag{4.13}$$

If grid-points are equally spaced in longitude λ and in $s = \sin\phi$ (rather than in ϕ), then $\mathbf{A}$ is a scalar multiple of the unit matrix; (4.6) reduces to the eigenvalue problem for $\mathbf{B}$, and the transformations (4.10) and (4.12) simplify. However, even in this case, the eigenvectors are not precisely equal to spherical harmonics evaluated at the grid-points.

5 Concluding remarks

In this paper we have noted that error covariance *matrix* eigenvalue problems arise as natural discretizations of spatially-continuous error covariance *integral* eigenvalue problems. When the error covariance is homogeneous and symmetric in a 1-dimensional, cyclic domain, the eigenfunctions in the continuous case are sinusoidal and the eigenvalues are the Fourier cosine amplitudes of the covariance profile. These results are well-known in various contexts; see, for example, Preisendorfer (1988), pp 156, 157. They are physically reasonable because the continuous covariance problem is isomorphic to a continuous smoothing problem, and it is a basic result in signal processing that applying any symmetric smoother to a periodic function gives the same periodic function but with a different amplitude. Indeed, this result is so basic that its implications can escape notice. When seeking the effect of a symmetric running smoother S on a waveform $\exp(ikx)$, i.e. when evaluating

$$R(k,\, X) \equiv \int S(x - X) \exp(ikx) dx, \tag{5.1}$$

one hardly pauses to suspect that $R(k,\, X)$ might not be of the form $r(k) \exp(ikX)$, where $r(k)$ is real function of k. The fact that $R(k,\, X)$ *is* of the form $r(k) \exp(ikX)$ means that $\cos kx$ and $\sin kx$ are eigenfunctions of the integral eigenvalue problem

$$\int S(x - X) f(x) dx = \mu f(X). \tag{5.2}$$

This result is a specialization of a convolution theorem for full-range Fourier series, but it is much easier to establish the result than the convolution theorem.

Discrete covariance eigenvalue problems may be considered to spatially sample continuous covariance eigenvalue problems. This makes clear why the eigenvectors for homogeneous, symmetric covariance are not sinusoidal in the index space of grid-points that are unevenly spaced; they remain so (or nearly so) when viewed in undistorted space. Analysis can proceed in such cases via the symmetric matrix $\mathbf{D}^{1/2}\mathbf{B}\mathbf{D}^{1/2}$, where $\mathbf{B}$ is the covariance matrix on the uneven grid and $\mathbf{D}$ is a diagonal matrix of grid separations (or, in 2 dimensions, of areas).

In cyclic domains in 2 and 3 Cartesian dimensions, sinusoidal eigenfunctions occur if the error covariance is separable into factors symmetric in each dimension. Such a separably-symmetric covariance is centrally-symmetric but not isotropic. Isotropy requires radial symmetry, which occurs in 2 or more cyclic Cartesian dimensions only in the trivial case of uniform covariance - a severe restriction to which the finite cyclic domain of a spherical surface is not subject.

On the sphere, and with homogeneous, isotropic error covariance, the eigenfunctions of the continuous eigenvalue problem are spherical harmonics, each eigenvalue being determined by the projection of the symmetric covariance function onto the Legendre polynomial (zonal harmonic) having the same degree as the corresponding spherical harmonic eigenfunction.

That spherical harmonics are eigenfunctions of the integral equation (4.1) implies their orthogonality, demonstrated by Boer (1983), p 158, in the "probability average" in homogeneous, isotropic turbulence on the sphere. Suppose that $f_1 = Y_{n_1}^{m_1}$ is an eigenfunction of (4.1) with eigenvalue μ_1, i.e. (in the notation of (4.5))

$$\int S(\omega - \Omega) f_1(\omega) d\omega = \mu_1 f_1(\Omega). \tag{5.3}$$

By multiplying (5.3) by the complex conjugate of another eigenfunction $f_2 = Y_{n_2}^{m_2}$ of (4.1) (i.e. $m_2 \neq m_1$ and/or $n_2 \neq n_1$), integrating over Ω and noting the orthogonality properties of the Y_n^m one obtains

$$\int\int S(\omega - \Omega) f_1(\omega) f_2^*(\Omega) d\omega d\Omega = 0. \tag{5.4}$$

Spherical harmonics are thus orthogonal with respect to the metric defined by the homogeneous isotropic covariance kernel $S(\omega - \Omega)$. According to Courtier *et al.* (1998), p1786, this is the "fundamental result at the heart of 3D-Var". Our focus in the present paper has been on the spherical harmonics as eigenfunctions of the covariance kernel $S(\omega - \Omega)$ itself, as expressed in the integral equation (4.5); this has led to a more transparent and more direct treatment for our present purposes.

The analysis is readily extended to the case of homogeneous *correlation* (which, as noted by Rutherford (1972), is a better idealization of forecast error behaviour). $\mathbf{B}$ is related to the matrix $\mathbf{C}$ of error correlation coefficients by

$$\mathbf{B} = \mathbf{VCV} \tag{5.5}$$

where $\mathbf{V}$ is the diagonal matrix of the square roots of the variances. Hence $\mathbf{B}^{-1} = \mathbf{V}^{-1}\mathbf{C}^{-1}\mathbf{V}^{-1}$, and in terms of the new variable $\mathbf{z} = \mathbf{V}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right)$ (1.2) becomes

$$J = J(\mathbf{z}) = \frac{1}{2}\mathbf{z}^T\mathbf{C}^{-1}\mathbf{z} + \frac{1}{2}\left(\mathbf{y} - \mathbf{y}^o\right)^T\left(\mathbf{E} + \mathbf{F}\right)^{-1}\left(\mathbf{y} - \mathbf{y}^o\right). \tag{5.6}$$

Since $\partial J/\partial \mathbf{x} = (\partial J/\partial \mathbf{z})\mathbf{V}^{-1}$ and $\mathbf{V}$ is readily invertible, the minimum of J with respect to $\mathbf{z}$ coincides with the minimum with respect to $\mathbf{x}$.

In the remaining discussion we shall refer to homogeneous and symmetric covariance and to sinusoidal eigenvectors, with the understanding that the cases of homogeneous and symmetric correlation coefficients and spherical harmonic eigenvectors on the sphere are covered too (and bearing in mind that symmetric covariance implies isotropy in 1D cyclic domains and on the sphere, but denotes separable symmetry in cyclic 2D or 3D Cartesian domains and does not imply isotropy).

The existence of sinusoidal eigenfunctions of $\mathbf{B}$ when covariance is homogeneous and symmetric means that the "background" quadratic form J^b in the cost function $J(\mathbf{x})$ (see (1.2)) may be diagonalised by transforming to new variables that are the Fourier coefficients of the fields $\mathbf{x}$. This result gives a useful sidelight on a scheme discussed by Parrish and Derber (1992). They state (p1749) that "An ideal choice of analysis variables would be the amplitudes of the eigenvectors of the background error covariance matrix specified in terms of the model variables. In this case the $\mathbf{B}$ matrix would then be diagonal with the eigenvalues of the background error covariance matrix in terms of the model variables along the diagonal." It is helpful to know that if homogeneous and symmetric covariance is assumed then the eigenvectors are sinusoidal, that the amplitudes are the Fourier coefficients of the untransformed fields, and that the eigenvalues are the Fourier coefficients of the covariance profile. Indeed, many data assimilation schemes work in terms of Fourier (or spherical harmonic) amplitudes, and involve assertions that amount to statements of homogeneous and symmetric error covariance; see, for example, Courtier *et al.* (1998), p1786, and Lorenc *et al.* (2000), p3000. In Part II of this study (White and Healy 2004) we investigate the

extent to which analytical treatment for homogeneous and symmetric covariance can usefully be applied in the full 3-dimensional, mixed cyclic/non-cyclic, spherical annulus domain of the Earth's atmosphere.

We have also shown that a transformation in which individual Fourier components are scaled by the square root of their corresponding eigenvalues gives a diagonal quadratic form of the J^b term in which all coefficients are unity. This is precisely the form achieved by square rooting the $\mathbf{B}$ matrix and applying a transformation of variables in order to improve the conditioning of the variational problem for the best estimate fields, but without assuming homogeneous and symmetric covariance (Lorenc 1992, Lorenc *et al.* 2000, p2994).

Our analysis also throws light on the conditioning problem by emphasising that the eigenvalues of homogeneous covariance eigen-problems are the Fourier (or spherical harmonic) amplitudes of the covariance profile. Since the range of eigenvalues of $\mathbf{B}$ is a major determining factor in the conditioning (see Lorenc *et al.* 2000, p2994), the Fourier transform of the covariance profile is clearly a key element. Scaling by the square root of the eigenvalues can improve the conditioning if all are non-zero (and positive), but it fails if a zero eigenvalue occurs, i.e. in the homogeneous case, if the covariance profile does not project onto one zonal harmonic (or more). The scaling also fails if a negative eigenvalue occurs, thus emphasising that the covariance profile should have a consistently positive Fourier spectrum. The relationship of the eigenvalues to the Fourier transform of the profile of homogeneous covariance also gives a direct physical interpretation of them, and has an immediate association with the smoothing analogy that has been discussed at length in earlier sections.

Acknowledgements

We thank G. Inverarity and N. B. Ingleby for their careful reading of and helpful comments on an earlier version of this paper. Useful conversations with P. Andrews, S. Clough, S. Dance, A. Lorenc and I. Roulstone are also gratefully acknowledged.

References

Boer, G.J. 1983 Homogeneous and isotropic turbulence on the sphere.
J. Atmos. Sci., **40**, *154-163*

Courant, R. and Hilbert, D. 1953 Methods of Mathematical Physics, Vol I. Interscience

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljević, D.,
Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M. 1998
The ECMWF implementation of three-dimensional variational assimilation
(3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.,* **124**, 1783-1807

Daley, R. 1991 Atmospheric data analysis. CUP

Gaspari, G. and Cohn, S. E. 1999 Construction of correlation functions in two and three

dimensions. *Q. J. R. Meteorol. Soc.,* **125**, 723-757

Healy, S.B. and White, A.A. 2004 Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Q. J. R. Meteorol. Soc.*, to appear

Ingleby, N.B. 2001 The statistical structure of forecast errors and its representation in the Met. Office global 3-D variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **127**, 209-231

Lorenc, A.C. 1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194

Lorenc, A.C. 1992 Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.*, **118**, 569-591

Lorenc, A.C., Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L.F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J. and Saunders, F.W. 2000 The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **126**, 2991-3012

Mathews, J. and Walker, R.L. 1965 Mathematical methods of physics. Benjamin

Parrish, D.F. and Derber, J.C. 1992 The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.,* **120**, 1747-1763

Preisendorfer, R.W. 1988 Principal component analysis in meteorology and oceanography. Elsevier

Ramsey, A.S. 1949 An introduction to the theory of Newtonian attraction. C.U.P.

Rodgers, C. D. 1990 Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.*, **95**, 5587-5592

Rutherford, I. 1972 Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.,* **29**, *809-815*

Sardeshmukh, P.D. and Hoskins, B.J. 1984 Spatial smoothing on the sphere. *Mon. Weather Rev.,* **112**, 2524-2529

White, A.A. and Healy, S.B. 2004 Error covariance matrices having sinusoidal eigenvectors. II: non-cyclic and mixed domains. Submitted to *Q. J. R. Meteorol. Soc.* (Available as *Forecasting Research Technical Report No. 447*, Met Office)

Wilkinson, J.H. 1965 The algebraic eigenvalue problem. Clarendon

Appendix: The eigen-problem of the 7×7 matrix (2.5)

Element B_{jk} of the matrix $\mathbf{B}$ given by (2.5) can be written as

$$B_{jk} = S_0 + \sum_{m=1}^{m=3} 2S_m \cos\left[m\left(k-j\right)\Delta\right], \tag{5.1}$$

in which $\Delta \equiv 2\pi/7$ and S_0, S_1, S_2 and S_3 are related to a, b, c and Δ by (5.8)-(5.11) (below). With $S_{-3} = S_3$, $S_{-2} = S_2$ and $S_{-1} = S_1$ (all real), (5.1) condenses to

$$B_{jk} = \sum_{m=-3}^{m=3} S_m \exp\left[im\left(k-j\right)\Delta\right]. \tag{5.2}$$

Multiply (5.2) by $\exp\left[-in\left(k-k_0\right)\Delta\right]$ (where $n = 0,\ 1,\ 2,\ 3$ and k_0 is a constant) and sum over k to obtain

$$\sum_{k=1}^{k=7} B_{jk} \exp\left[-in\left(k-k_0\right)\Delta\right] = \sum_{m=-3}^{m=3} S_m \exp\left[-i\left(mj-nk_0\right)\Delta\right] \sum_{k=1}^{k=7} \exp\left[i(m-n)k\Delta\right]. \tag{5.3}$$

The sum over k on the right-hand side of (5.3) is equal to $7\delta_{mn}$, where $\delta_{mn} = 1$ if $m = n$ but $\delta_{mn} = 0$ if $m \neq n$. Hence (5.3) reduces to

$$\sum_{k=1}^{k=7} B_{jk} \exp\left[-in\left(k-k_0\right)\Delta\right] = 7S_n \exp\left[-in\left(j-k_0\right)\Delta\right]. \tag{5.4}$$

Thus the column vector $\mathbf{e}^n$ whose kth element is $e_k^n = \exp\left[-in\left(k-k_0\right)\Delta\right]$ is an eigenfunction of $\mathbf{B}$ with eigenvalue $7S_n$. But the elements of $\mathbf{B}$ are all real, so both the real and the imaginary parts of $\mathbf{e}^n$ are individually eigenfunctions of $\mathbf{B}$ with the same eigenvalue $7S_n$. In other words, the column vectors $\mathbf{c}^n$ and $\mathbf{s}^n$ whose kth elements are respectively $\cos\left[n\left(k-k_0\right)\Delta\right]$ and $\sin\left[n\left(k-k_0\right)\Delta\right]$ are eigenfunctions of $\mathbf{B}$ with eigenvalues $7S_n$:

$$\sum_{k=1}^{k=7} B_{jk} \cos\left[n\left(k-k_0\right)\Delta\right] = 7S_n \cos\left[n\left(j-k_0\right)\Delta\right],\ \ n = 0, 1, 2, 3\,; \tag{5.5}$$

$$\sum_{k=1}^{k=7} B_{jk} \sin\left[n\left(k-k_0\right)\Delta\right] = 7S_n \sin\left[n\left(j-k_0\right)\Delta\right],\ \ n = 1, 2, 3\,. \tag{5.6}$$

These are the required results. To relate S_0, S_1, S_2 and S_3 to a, b, c and Δ, the most direct procedure, as pointed out to us by Dr G. Inverarity, is to take the real part of (5.4) and set $k_0 = j$:

$$7S_n = \sum_{k=1}^{k=7} B_{jk} \cos\left[n\left(k-j\right)\Delta\right],\ \ n = 0, 1, 2, 3 \tag{5.7}$$

Upon recalling that (for example) $B_{41} = B_{47} = c$, $B_{42} = B_{46} = b$, $B_{43} = B_{45} = a$, $B_{44} = 1$, (5.7) gives (in terms of $\Delta \equiv 2\pi/7$):

$$7S_0 = 1 \ + \ 2a \ + \ 2b \ + \ 2c\,; \tag{5.8}$$

$$7S_1 = 1 \ + \ 2a\cos\Delta \ + \ 2b\cos 2\Delta \ + \ 2c\cos 3\Delta\,; \tag{5.9}$$

$$7S_2 = 1 \ + \ 2c\cos\Delta \ + \ 2a\cos 2\Delta \ + \ 2b\cos 3\Delta\,; \tag{5.10}$$

$$7S_3 = 1 \ + \ 2b\cos\Delta \ + \ 2c\cos 2\Delta \ + \ 2a\cos 3\Delta\,. \tag{5.11}$$

.

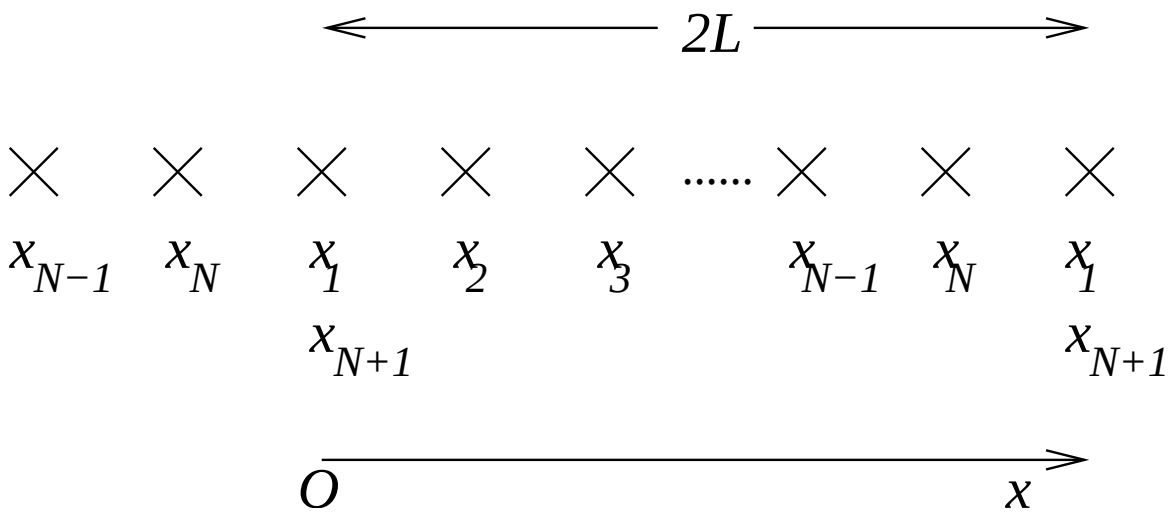

Figure 1: Equally-spaced grid-points x_1, x_2,.., x_N, x_{N+1} in a cyclic domain of length $2L$.

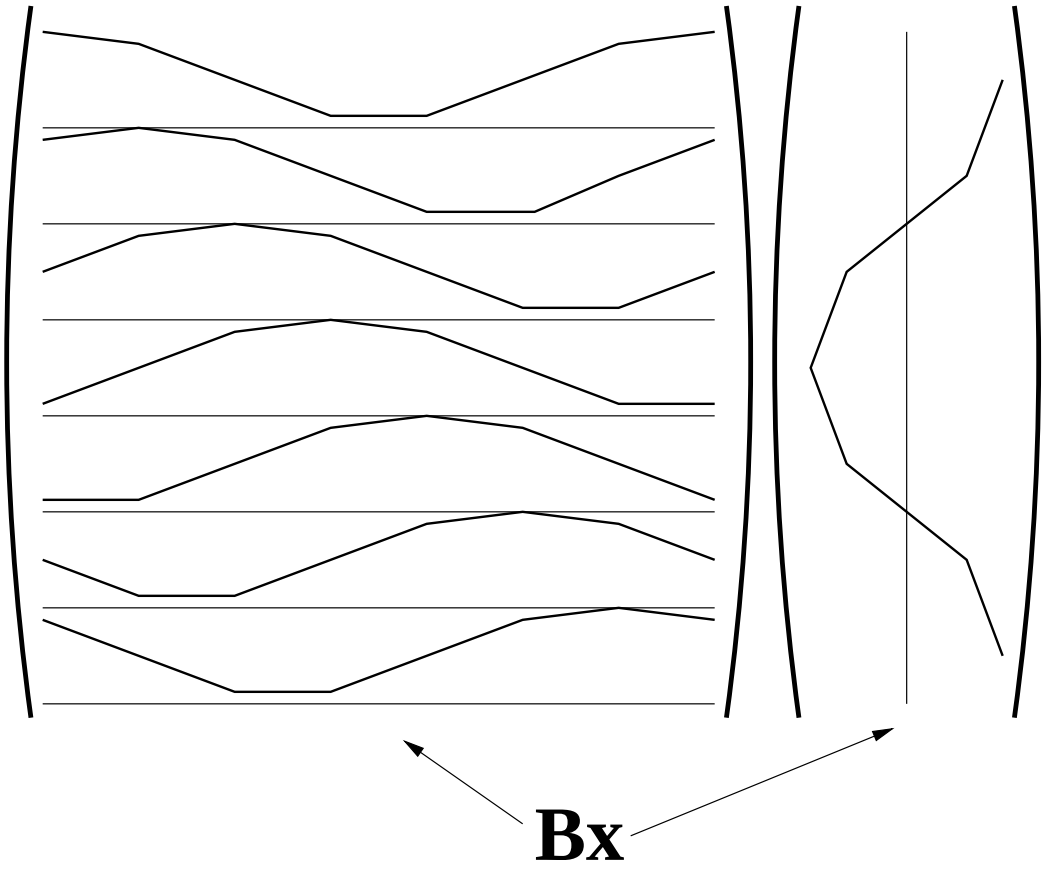

Figure 2: Depiction of a 7×7 circulant matrix $\mathbf{B}$ acting on a column vector $\mathbf{x}$ of wave 1 cosine form. The 7 elements of each row of $\mathbf{B}$ are plotted vertically and the 7 elements of $\mathbf{x}$ horizontally. Each row of the matrix has the same profile, but a uniform shift from left to right occurs with each succeeding row. The action of $\mathbf{B}$ on $\mathbf{x}$ will leave contributions only from the Fourier components of the rows of $\mathbf{B}$ which project onto the wave 1 cosine, and the contribution will change progressively with row number as the left-to-right shift of the profile occurs. Detailed analysis (see Appendix) shows that the operation $\mathbf{Bx}$ generates $\mathbf{x}$ multiplied by the coefficient of the wave 1 cosine in the Fourier decomposition of the symmetric row profile. Similar results hold if $\mathbf{x}$ is a wave 0, 2 or 3 cosine or a wave 1, 2 or 3 sine.

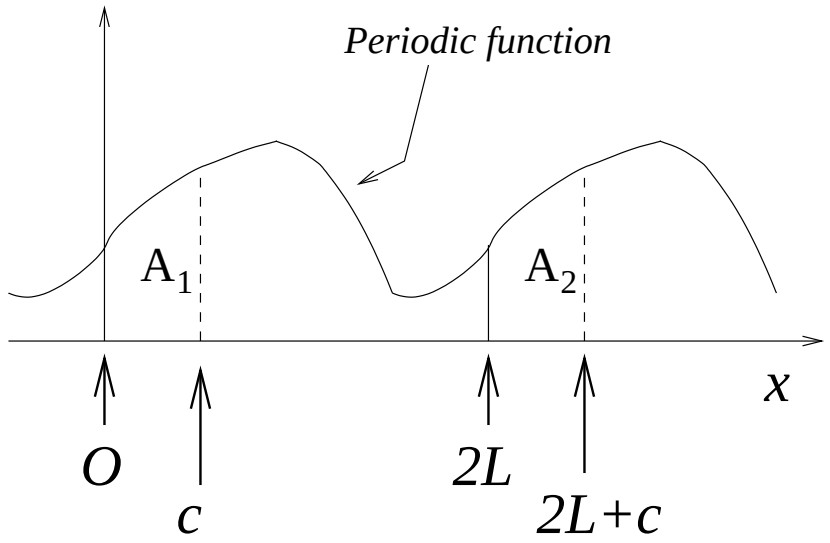

Figure 3: Because areas A_1 and A_2 are equal, the integral of a periodic function over one cycle (length $2L$) beginning at $x = 0$ is equal to the integral over one cycle beginning at $x = c$.

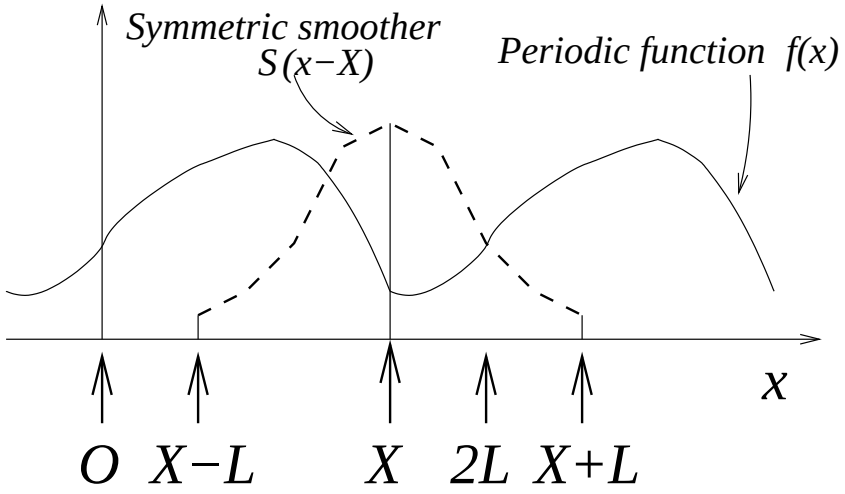

Figure 4: A periodic function $f(x)$ may be smoothed by replacing its value at $x = X$, i.e. $f(X)$, by the result of integrating the product $S(x - X)f(x)$ over the interval $X - L \leq x \leq X + L$. This is equivalent to placing the origin of x at X, and replacing $f(X)$ by the integral of the product $S(x)f(x + X)$ over the interval $-L \leq x \leq L$.

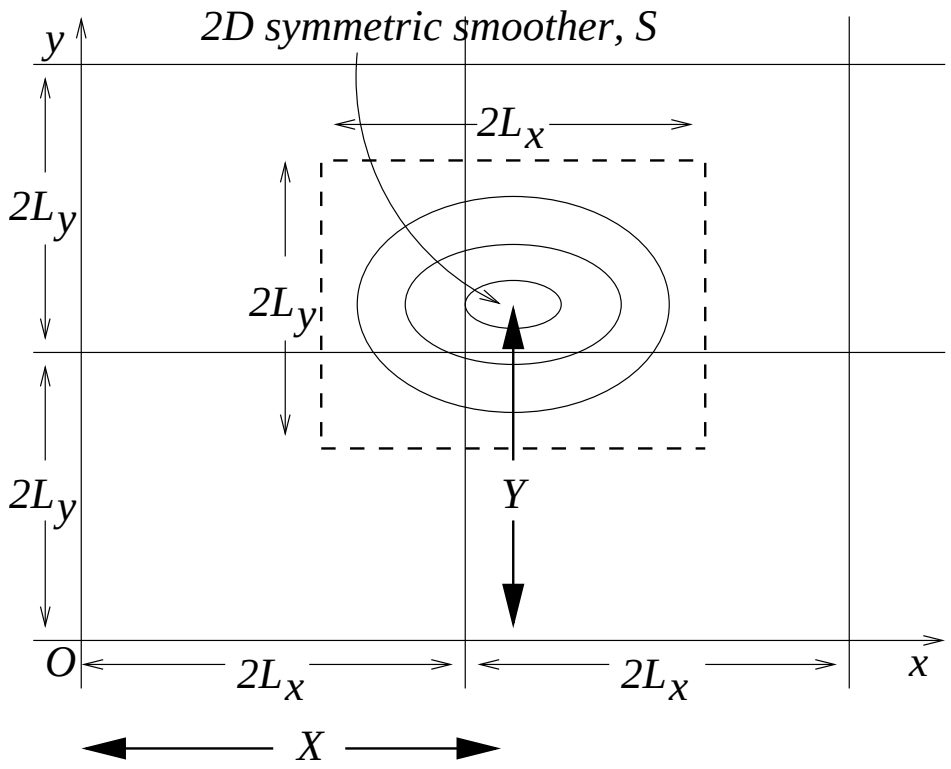

Figure 5: Part of a 2D cyclic domain having repeat distances $2L_x$ and $2L_y$ in x and y. Also shown are contours of a centrally-symmetric smoother $S(x-X,\ y-Y)$ which, when applied to any function f periodic in the domain, replaces $f(X,Y)$ by the integral of the product $S(x-X,\ y-Y)f(x,y)$ over the rectangle $X-L_x \le x \le X+L_x$, $Y-L_y \le y \le Y+L_y$. [The contours of S have been drawn as ellipses for convenience in drafting. A smoother of precisely this form cannot be factored as in (3.2), and so does not have the property of separable symmetry which we have assumed in demonstrating the existence of sinusoidal eigenfunctions.]

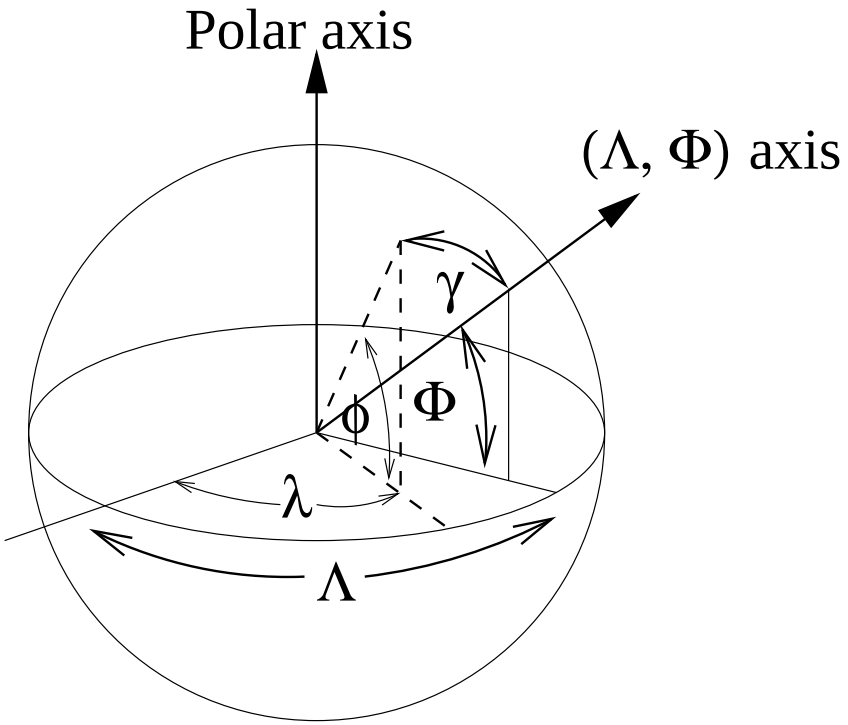

Figure 6: Showing the longitudes λ and Λ and the latitudes ϕ and Φ of two points on a sphere, and the angle γ subtended at its centre by their radii . The geographical polar axis and the axis defined by the centre of the sphere and the point (Λ, Φ) are also shown.